YORK
Health Economics
CONSORTIUM

MET OFFICE

Review of the Met Office Model for Emergency Admissions to Hospital

Final Report

PAUL SCUFFHAM, Senior Research Fellow
ADRIAN BAGUST, Project Director Disease Modelling

APRIL 2003

INVESTOR IN PEOPLE

Market Square, University of York, Vanbrugh Way, Heslington, York YO10 5NH
Tel: 01904 433620 Fax: 01904 433628 Email: yhec@york.ac.uk http://www.york.ac.uk/inst/yhec/

THE UNIVERSITY of York
York Health Economics Consortium is a Limited Company
Registered in England and Wales No. 4144762 Registered office as above.

CONSORTIUM

Contents

Executive Summary

1. BACKGROUND AND OBJECTIVES

Statistical models were developed by the Southampton group (led by Arjan Shahani) to forecast emergency admissions to hospitals on a regional basis for the Met Office's *Forecasting the Nations Health* (FNH) project.

There were 156 models developed to forecast emergency admissions for:

- Medical;
- Non-medical;
- 0-14 years;
- 15-64 years;
- >64 years;
- Total emergency admissions.

This report identifies, describes and critically appraises the technical characteristics of these models as part of YHEC Ltd's evaluation of FNH. The focus of this review is on:

- Data used to model emergency admissions;
- Estimation techniques used to project NHS activity;
- Statistical diagnostic tests;
- Future areas for development/improvement.

2. METHODS

We comment here on the census data used to construct 5 socioeconomic clusters for England, the construction of the temperature variables for each of the socioeconomic clusters, the other explanatory variables, and the methods for estimation.

Diagnostic tests of model adequacy are conducted using four models (two *good* and two *poor*) from two regions:

- Northern Region, socioeconomic group 1, (i.e. the *poor* models);
- Southeast Region, socioeconomic group 1, (i.e. the *good* models).

For each of these regions, models for total admissions and medical admissions were assessed.

Diagnostic tests applied included tests for autocorrelated errors, heteroskedasticity (i.e. non-constant variance), normality of residuals (i.e. whether the error had the properties of a normal distribution), non-stationarity (i.e. trends in the error terms), and independence (i.e. correlation of errors between models).

3. RESULTS

There are nine Public Meteorological Service (PMS) regions in England. These were disaggregated into 26 postcode clusters based on socioeconomic characteristics. Up to five socioeconomic groups were identified within each of these PMS regions, through a principal components analysis and a cluster analysis was a useful and a commendable approach, facilitating detailed cross-section data to be used in time-series analysis. The five socioeconomic groups were later aggregated into three groups; this has led to some inefficiency (statistically) with respect to the weather variables. In addition, the census data provided, on which this socioeconomic grouping was performed, was restrictive and is inevitably likely to be out of date despite it being the most up-to-date data available at the time.

Isothermal cluster analysis was also undertaken to identify different areas within each of the five socioeconomic groups, within each PMS region. Weather data was restricted to maximum temperature and lowest maximum temperature for each of the isothermal clusters in each socioeconomic group, for the previous (arbitrarily chosen) 2, 7, 10 and 15 days. The aggregation of socioeconomic groups 3, 4 and 5 into one group meant that temperature data from one isothermal region in one socioeconomic group is potentially used to forecast emergency admissions in another socioeconomic group. The clustering of weather data into isothermal clusters appears to be of limited value, it is cumbersome to work with, and offers little explanatory power. An improved approach might be to use weather data for each socioeconomic group as a whole rather than disaggregating into isothermal clusters.

Models were developed using daily data on hospital admissions, weather and GP consultation rates for infectious diseases for the four years from April 1995 to March 1999. A step-wise regression procedure was used to include/retain/exclude variables from the models. The coefficients from these models are used to forecast admissions in 2003. This is four years later, and it is not unlikely that socioeconomic boundaries and management practices for emergency admission have changed. In addition, it is likely that there may be substantial population changes in some areas over this time, with increases in total population and the proportion of older people.

With the exception of simple correlation coefficients between the forecast and actual admissions, no statistical diagnostic tests were undertaken by Southampton. The statistical tests undertaken by YHEC show there is substantial autocorrelation in errors. There was some heteroskedasticity in one of the four models tested. Tests for trends showed substantial non-stationarity in three of the four models. One model had errors that were not normally distributed, and all models suffered from either excess kurtosis or skewness. In addition, there was statistically significant correlation of errors between regions, suggesting that factors causing forecast errors in one region also cause forecast errors in another (i.e. non-independence of errors). Given these diagnostics, variables are likely to have been incorrectly included or excluded in the step-wise procedure used.

Implementing the models by the Met Office has required an error-correction mechanism (“error tracking”) to adjust forecasts to reasonable levels. It is noted that this was the first attempt at developing a model and implementing the model with live forecasts. As such, suggested improvements will be beneficial. As a minimum, trends should be included in the models, and ideally, if the models are to be used for the next few years, short-run and long-run dynamics should be incorporated.

4. DISCUSSION

The models developed by Southampton were a major undertaking and form the basis for future development. Many aspects of the model development are commendable (for example, the socioeconomic data from the 1991 Census was used in such a way that enabled geographic regions within England to be defined and hospital admission data to be subsequently disaggregated; see Section 2.1).

The models developed for total admissions provided the best fit between actual and forecast admissions (with a mean goodness-of-fit of 71%) and models for the 0-14 year age-group provided the worst fit (mean goodness-of-fit of 34%). That is, forecasting accuracy was better with higher levels of aggregation. However, modelling the volatility apparent in disaggregated data is of greatest use and interest.

There are many limitations with the models developed; some of these are relatively minor and pertain to the data used, some are statistical issues on how the data was used, and some are conceptual issues. All of these limitations need to be addressed in future model-development work on forecasting hospital admissions.

The major limitations were:

- Socioeconomic clusters were based on the 1991 census; these groupings are likely to have changed over the last 12 years and/or compared with the 2001 census;
- The models were constructed with hospital and temperature data from April 1995 to March 1999; the coefficients from that period were used to generate forecasts in 2003. Forecasts made four years into the future based on four previous years of data (although this was the best available at the time) are unlikely to be robust. (This is the subject considered in a separate report by YHEC reviewing the forecasts for the winter of 2002/03);
- Trends in hospital admissions and population growth were not considered. This results in poor forecasts and invalid statistical tests (e.g. *t*-tests and *F*-tests).

There are some important aspects that need to be refined and addressed in future development of the models. These include:

- Updating socioeconomic cluster boundaries based on the 2001 Census data;
- Aggregating temperature variables to match with the boundaries of the socioeconomic groups;
- Including trends in the analysis;
- Avoiding use of a step-wise approach or control for autocorrelation;
- Including dynamics for population and hospital changes over time.

Acknowledgements

The authors gratefully acknowledge the help from Dr Arjan K Shahani and Adam Brentnall (Faculty of Mathematical Studies, University of Southampton) in providing the data, models, the detailed S-Plus printouts of results, tables of correlation coefficients, and the helpful Excel spreadsheets for evaluating the forecasts.

In addition, the authors are grateful for the comments received from other members of the YHEC Ltd team on the previous draft of this report.

Section 1: Background

1.1 BACKGROUND

As part of the *Forecasting the Nation's Health* (FNH) project, the Met Office commissioned the University of Southampton to develop statistical models to forecast emergency admissions to hospital. For medical emergencies dependent on weather conditions (temperature) and infectious diseases, forecasts out to nine days ahead and distributions of length of stay were the key variables of interest. It is anticipated that hospital managers can adjust patient loads and staff workloads according to the expected changes in emergency admissions.

YHEC Ltd was commissioned to evaluate the FNH project. There are numerous different approaches to modelling the association between weather patterns and admissions to hospital, and the Southampton group have chosen an approach that they consider most suitable. In this report we offer comment on the validity of the data used, the overall design of the models and approach to modelling, and, using historic data for testing the models, how well the models actually forecast emergency admissions. This last point is the litmus test and is the subject of a separate validation process.[1]

1.2 SUMMARY DESCRIPTION OF THE MODEL(S)

A large set of independent models was developed to forecast the numbers of emergency admissions for hospitals in England. Additional statistics were performed on the distribution of length of stay to provide static tables for each Trust, but to-date the YHEC team have not seen documentation or forecasts for this component.

Statistical models were developed using postcode-level data. Because of the large number of postcodes in England (approximately 2000), postcodes were grouped into regions with similar characteristics, based on weather profiles and socioeconomic characteristics. This grouping process was an important feature early in the modelling process. In addition, exploratory analyses (CART) was undertaken to determine the potential importance of each variable in explaining emergency admissions. Each of the nine Public Meteorological Service (PMS) regions in England were divided into two to three socioeconomic regions according to 1991 census data. Then, each socioeconomic region was categorised according

[1] We are undertaking a separate review of the accuracy and reliability of the forecasts from the models. For this we are using the data supplied to us from The Met Office on the forecast and actual numbers of emergency admissions for the winter of 2002/03.

to three-levels of iso-thermal profiles. Overall, there are 26 basic models (9 PMS regions x 3 socioeconomic subdivisions).

Final models were developed to explain and forecast emergency admissions. The dependant variables were the numbers of emergency admissions separated into:

- Medical;
- Non-medical;
- 0-14 years;
- 15-64 years;
- >64 years;
- Total.

Therefore, there are a total of 156 (6 x 26) models.

Seasonal and typical weekly fluctuations were modelled by using dummy variables for "day of week", and "month of year". Variables relating to temperature and infectious diseases (the latter obtained from the Royal College of General Practitioners infectious disease surveillance system) were included as continuous exogenous variables. A step-wise regression procedure that retained variables that were significant at the 5% level was used to select variables and estimate parameters. The explanatory power of the models showed a goodness of fit between forecast and actual admissions ranging from 0.58 to 0.85 for total admissions. There were a large number of observations (four years of daily data: n = 1,461) and, therefore, the goodness-of-fit could be expected to be better; however, four years is a relatively short horizon in time series analysis. Goodness-of-fit deteriorated with disaggregation of admissions into medical/non-medical and was poorest in the youngest age groups (0-14 years). Based on these variables, forecasts up to nine days ahead were made.

For hospitals reporting actual numbers of emergency admissions, a crude updating algorithm, added to the forecasts by the Met Office, was used in an attempt to correct for forecast errors.

The models developed were static models – that is, dynamics are not fed back in to update and revise the parameter estimates. Long-run trends and cycles are not incorporated in the models.

1.3 OBJECTIVE OF REVIEW

The aim is to identify and describe the technical characteristics of the models and the links between weather and utilisation of services. The focus of this review is on:

- Data used to model emergency admissions;

- Estimation techniques used to project NHS activity;
- Statistical diagnostic tests; and
- Future areas for development/improvement.

The primary documents used for this review are those supplied from the Met Office, which in turn were provided by the Southampton group detailing the development of the models. The key documents are:

i. Development of postcode level models for chosen types of daily emergency admissions to hospitals in England, and, Preparation of Tables that summarise the distributions of lengths of stay for chosen medical and non-medical conditions

Author: Dr Arjan Shahani, University of Southampton
Date: 09 December 2002

ii. Model of Hospital Workload, MODEL DEVELOPMENT GUIDE, Version: 1

Author: Clare Bryden, Civil Applications, Met Office
Date: 12 December 2002

iii. Description of Model (on Website) Forecasting the Nation's Health – Model of Hospital Workload

Author: The Met Office
Date: 06 December 2002

These three documents are hereafter referred to as Doc 1, Doc 2 and Doc 3 respectively. In addition, a sample of the data, the models, the detailed S-Plus printouts of results, and the Excel spreadsheets for evaluating the forecasts (supplied by Southampton) were used.

The remainder of this report is structured in three sections. The next section (Section 2) describes the data, variable selection process, emergency admission model construction and the explanatory power of the models and diagnostic tests. Section 3 describes implementation of the models by the Met Office focussing on modifications to the models. The final section discusses the overall findings including the general conclusions of the review and suggestions for developments and future research are made.

Section 2: Review of the Model(s)

2.1 DATA USED AND METHODS OF DATA AGGREGATION

Census data

A large body of literature has developed showing that socioeconomic factors are important determinants of admission to hospital.[2] However, potentially more important factors are demographic factors, such as the average age of the population, and especially the proportions of older people and the very young who have higher admission rates to hospital and consume more health care resources. The 1991 Census captures both socioeconomic and demographic factors and this was used to develop socioeconomic profiles for postcode areas. The Met Office suggested 66 variables to use from that Census.

The Met Office typically separates England into nine Public Meteorological Service regions (PMS regions). For each of these nine regions, a principal components analysis (PCA) was used to generate linear combinations of these Census variables. Twenty-five variables, in five linear combinations, captured 87% of the variation in the data. Thus, Southampton concluded that postcodes for England should be grouped in five socioeconomic classes using these 25 variables (Table 2.1).

[2] For example, see: Jarman B (1984), Kuh D, Stirling S (1995), Duffield JS, Craig K, Plant WD (1997), Reid FD, Cook DG, Majeed A (1999), Bennett J, Pitman R, Jarman B, et al. (2001).

Table 2.1: 1991 Census data

Category of variable	Levels of variable (from Census)	Aggregated levels of variable	Comments
Children at home (1 vs 2 or more)	5	4	H/holds with children of mixed (0-4 and 5-15 years) omitted
Employment status	10	8	Retired M&F excluded
Ethnic origin	10	1	White only included
Household size (persons)	8	1	1 person only H/hold
House size (rooms)	2	2	
Marital status	4	2	Single, widowed, divorced M&F excluded
Mobility (car vs no car)	1	1	
Socioeconomic group (work place and skills)	16	3	Grouped into groups 1-4, 5-6, 7-11 (12+ excluded)
Transport (to work)	8	3	Train/bus, car other
Rejected variables			
Housing type (detached house…)	9		
Lifestage (i.e. lone parent)	4		

Following the PCA, cluster analysis was carried out to create the five groups (socioeconomic clusters 1-5); however, Southampton later decided (due to small sample sizes in some postcodes) that clusters 3-5 should be grouped together in all PMS regions, and in Devon and Cornwall, groups 1 and 2 should be grouped together for model estimation only (Table 2.2). However, the five clusters were retained for constructing and assigning weather variables (see below).

Table 2.2: Socioeconomic clusters within PMS regions

PMS Region	Socioeconomic clusters		
Central Southern	1	2	3,4,5
Devon & Cornwall	1,2		3,4,5
Eastern	1	2	3,4,5
Midlands	1	2	3,4,5
North West	1	2	3,4,5
Northern	1	2	3,4,5
South East	1	2	3,4,5
West	1	2	3,4,5
Yorkshire & Lincolnshire	1	2	3,4,5

This Census data is less than ideal. The Census data is now almost twelve years old. Due to changes in demographics, including an ageing population and urban sprawl, this data is possibly out of date, and hence, may be of limited value (however, we acknowledge that many agencies are forced to rely on it and at the time it was the best available).

The list of 66 variables supplied to Southampton for selection was potentially restrictive; however, working with all possible variables would be cumbersome, and in a project already saturated with data, more variables might detract from the main objective at hand. The 1991 Census household data contains 120 variables (of which 10% are identifying variables such as "Ward code") and the 1991 Census individual data contains 67 variables.

Any list of variables related to health and admissions to hospital should include:[3]

- The total population (or household size so that an estimate of the population can be calculated);
- The proportion of elderly;
- The proportion of children;
- Type of house (as houses with stairs can have implications for falls/injuries);
- Mode of transport to work (as traffic crashes are directly linked to emergency admissions).

In addition, other factors might include:

- (Un)employment status;
- Highest educational qualification;
- Tenure of housing;
- Socioeconomic group and/or ethnic origin.

The contribution of these factors to explaining admissions to hospital is, in itself, a topic of interest but, as yet, their role is undetermined.

There are some minor inconsistencies in the documentation on the variables used to identify the socioeconomic groups. For example, "housing type (detached, semi-detached, etc.) appears in the list of chosen variables but not in the factor analysis; conversely, house size (number of rooms) appears in the PCA but not in the list of chosen variables. Similarly, Southampton refer to Social Class (Doc 1), whereas it was socioeconomic group (SEG) that was suggested by the Met Office and used by Southampton. These are minor points where care is needed.

The overall approach is commendable as it enables cross-section data to be used with the time-series. That is, England was separated in to the nine PMS regions, and within each of these regions, sub-areas were identified according to socioeconomic characteristics. Time-series data for each of these sub-areas was then used for forecasting admissions. This approach should be maintained in future analyses, but updated with the recent Census data as it becomes available.

[3] Some of these were included either directly or in an aggregated form.

Weather data

CART, PCA and Cluster analyses were also carried out on 1999 weather data for daily minimum, mean and maximum temperatures within each of the nine PMS regions. This resulted in the grouping of postcodes into isothermal sectors within each of the 5 socioeconomic areas within each PMS region. The number of isothermal sectors varied between 2-4. However, disaggregating socioeconomic areas into isothermal areas would result in 486 models to estimate. This would have been inefficient. Therefore, weather variables (temperature) were subsequently used as explanatory factors in the regression analyses.

Minimum temperatures were found during the CART analysis and Principal Components Analysis to be less important as they generally occur during the night (Doc 3). Therefore, minimum temperatures were excluded from further analysis. However, minimum temperatures may be an important factor as they could indicate freezing; freezing is associated with more hazardous travel (walking or driving). A dichotomous variable for *freezing* could prove useful. Similarly, exploration of *precipitation* could prove useful.

The independent weather variables used in the final models were:

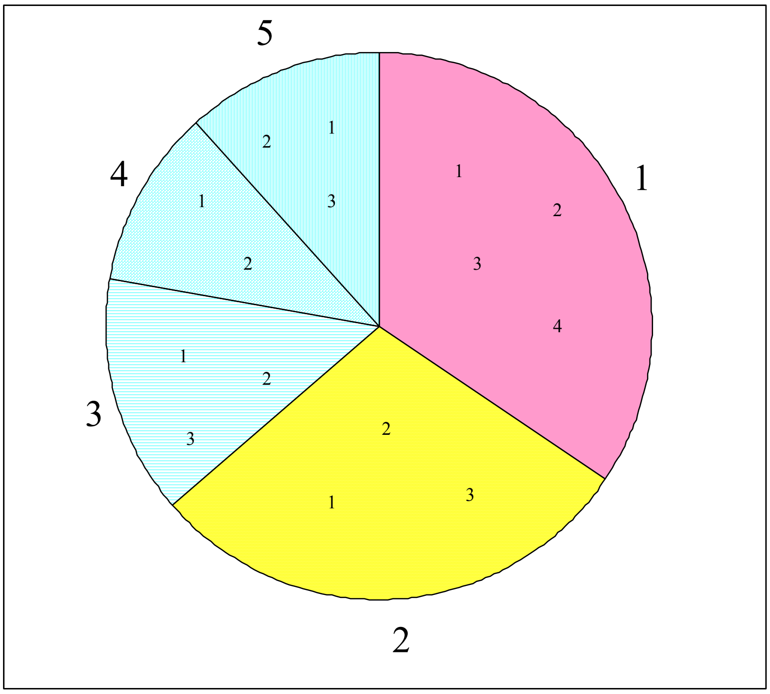
- Maximum daily temperatures 2, 7, 10 and 15 days ago, averaged over the postcode districts in each socioeconomic cluster and isothermal region;

- The lowest (minimum) of maximum temperatures over the periods from 7, 10 and 15 days ago to the current day, averaged over postcode districts.

This created seven primary temperature variables for inclusion as exogenous factors in the forecasting regression models. It was decided that no other weather variables should be included at this, the first stage of the model development. As these weather variables were developed at a stage when the socioeconomic groups were separated into five clusters (within each PMS), and each socioeconomic area was disaggregated into two to four isothermal areas, there were up to 28 weather variables for a single socioeconomic area (i.e. 7 x 4). In groups where the socioeconomic areas were aggregated (i.e. socioeconomic groups 3-5) there were three times as many weather variables (i.e. 7 weather variables for each isothermal area within each socioeconomic area: 7 x 3 x 3 = 63). (An example of the list variables used is given in Appendix A.)

Disaggregating weather into the isothermal regions seems to have added little value, and means that forecasts of admissions in one region were potentially based on the weather in another region. Because socioeconomic groups 3-5 were later aggregated, the temperatures for isothermal sector X in socioeconomic group Y is used as a potential explanatory factor to forecast admissions across the whole socioeconomic group 3-5. Figure 2.1 is used to illustrate this point. Each of the 5 socioeconomic groups is represented with the large digits around the outside of the figure, and the numbers 1-4 within each socioeconomic group represents each isothermal sector. Isothermal sector 3 in socioeconomic group 5 is used as an explanatory variable to forecast admissions for the combined socioeconomic groups 3, 4 and 5. Isothermal sector 3 in socioeconomic group 5 might have little relationship whatsoever with admissions to hospital in socioeconomic group 3.

Figure 2.1: An illustrative example of isothermal sectors within socioeconomic groups within a PMS region



A more efficient (and less cumbersome) approach would have been to calculate the seven primary weather variables for each socioeconomic area.

The decision to limit temperature data to days 2, 7, 10 and 15 days ago was due to a combination of both medical knowledge and the statistical evidence. High collinearity exists between daily temperatures; however, based on the step-wise regression procedure used it would have been relatively easy to include temperatures for all of the past 15 days – only those that were significant would be retained. This is a relatively minor limitation because the chosen weather factors had only a small contribution to forecasts (see below).

Data on other explanatory factors

In all models, dummy variables were included for month of year and day of week (plus a dummy variable representing public holidays). These dummy variables were the most important factors in the models. However, no explanatory factors were included to capture trends over time, or cycles.

Trends in admissions to hospital arise from factors such as changes in population and age-structure of the population. For example, over the three years 1998-2000, there was an average growth in population of 0.4% annually; in the 0-14 age-group population growth was negative (–0.2%) over the same period (and was –0.4% last year) but in the elderly (80+ years) there was an average growth in population of 0.6% annually (with 1.1% growth last year). These older persons are more likely than average to be admitted to hospital, and therefore, these increases in population should be taken into account to reduce the potential for systematic under-estimation of admissions.

Estimates of the GP consultation rates, lagged one week, for influenza and bronchitis in the postcode districts in each socioeconomic cluster and isothermal region were used as an additional explanatory factor. GP consultation rates were obtained from the weekly returns service (WRS) of the Birmingham Research Unit of the Royal College of General Practitioners (RCGP). This service involves 78 practices in England and Wales covering a population of 506,782. GPs report weekly numbers of consultations for influenza-like illness (ILI) and various other infectious and respiratory conditions including bronchitis plus some major medical conditions (acute myocardial infarcts (AMI), cerebral vascular accidents, (CVA)). These GP consultation rates are potentially an important explanatory variable as respiratory syncytial virus (RSV), mycoplasma pneumoniae, bronchitis and influenza viruses (classified into subtypes A, H1, H3, B and N1, N2) have been shown to be associated with admissions to hospital.[4] In future studies/analysis, the consultation rate data could be extended to include consultations for RSV and mycoplasma pneumoniae, and possibly AMIs and CVAs.

2.2 MODEL ESTIMATION

Following the exploratory analyses (above) where clusters and variables were defined and selected, 156 models were developed with daily observation data from 1 April 1995 to 31 March 1999. This gave four years of daily observations (1,461 observations for each socioeconomic cluster in each PMS region). Data were untransformed and models developed were strictly linear.

The models are represented by an equation of the form:

$$Y_{i,t} = a_0 + b_{j1}\,MaxT_{j,t} + b_{j2}\,MinMax_{j,t} + b_R\,RCGP + b_M\,Month + b_D\,Day + error_t$$

4 Fleming DM (2000), Fleming DM, Cross KW, (1993), Nguyen-Van-Tam JS, Brockway CR, Pearson JCG, et al. (2001), Scuffham P (2003), Simonsen L, Fukuda K, Schonberger LB, Cox NJ, (2000)

- Y_{it} = admissions for socioeconomic group *i* at time *t*;
- a = constant;
- i = socioeconomic sector {1,2,3-5};
- j = socioeconomic sector {1,2,3,4,5}
- b_x = parameters to estimate {x = j1 j2, R, M, D}; R= RCGP, M=Month, D= Day
- MaxT = Maximum daily temperature for the j^{th} socioeconomic group {k = previous 2, 7, 10 and 15 days};
- MinMax = lowest of the maximum temperatures for the previous 7, 10 and 15^{th} days for the j^{th} socioeconomic group;
- RCGP = Royal College of General Practitioners consultation rates for infectious diseases;
- Month = dummy variable for month of year {2,…12; month 1 = numeraire};
- Day = dummy variable for day of week {2,3,4,5,6,7,8; day 1 = numeraire; day 8 = public holiday}; and
- error = estimation error.

S-Plus was used to estimate the parameter coefficients for the 156 models. A step-wise regression procedure was used throughout. The dummy variables (day of week and month of year) and the exogenous variables (i.e. the temperature and RCGP variables) were sequentially entered and removed until a parsimonious model was obtained. A 5% level of significance was used as the retention criteria for the exogenous variables. All models retained the day of week variables, but in some models the month of year variables were rejected. Rejection on the month of year indicates there was no significant seasonality around a monthly cycle phase.

To test the reliability of the models, the coefficients were used to forecast daily admissions from 1 April 1999 to 31 March 2000. (In preliminary analyses, the RCGP variable was excluded; however, the models with RCGP started running 'off-line' by the Met Office in February 2003.) The correlation coefficients between the forecast and actual admissions for the 156 models are shown in Table 2.3.[5]

[5] There are *r* values and not the goodness of fit R^2 measure.

Table 2.3: Correlation coefficients of forecasts vs actual admissions for 1999 (excluding RCGP)

	Total	Medical	Non-medical	Age 0-14	Age 15-64	Age 65+
Central Southern SE1	0.77	0.74	0.61	0.38	0.61	0.74
Central Southern SE2	0.70	0.67	0.45	0.30	0.48	0.68
Central Southern SE3-5	0.73	0.69	0.46	0.26	0.46	0.71
Devon & Cornwall SE12	0.59	0.54	0.35	0.12	0.43	0.52
Devon & Cornwall SE3-5	0.66	0.63	0.46	0.20	0.47	0.65
Eastern SE1	0.80	0.77	0.70	0.49	0.70	0.78
Eastern SE2	0.70	0.66	0.61	0.35	0.60	0.69
Eastern SE3-5	0.57	0.53	0.44	0.25	0.43	0.51
Midlands SE1	0.73	0.73	0.53	0.37	0.65	0.72
Midlands SE2	0.80	0.78	0.61	0.49	0.76	0.79
Midlands SE3-5	0.72	0.70	0.52	0.33	0.61	0.67
North West SE1	0.76	0.73	0.59	0.39	0.61	0.69
North West SE2	0.81	0.78	0.69	0.50	0.75	0.73
North West SE3-5	0.76	0.71	0.60	0.37	0.65	0.67
Northern SE1	0.52	0.49	0.25	0.17	0.33	0.37
Northern SE2	0.73	0.72	0.58	0.35	0.63	0.66
Northern SE3-5	0.68	0.65	0.46	0.28	0.49	0.59
Southeast SE1	0.85	0.82	0.74	0.45	0.76	0.80
Southeast SE2	0.79	0.75	0.63	0.35	0.65	0.72
Southeast SE3-5	0.87	0.83	0.80	0.54	0.84	0.78
West SE1	0.61	0.62	0.37	0.24	0.40	0.59
West SE2	0.58	0.55	0.41	0.21	0.44	0.51
West SE3-5	0.60	0.58	0.41	0.19	0.41	0.57
Yorkshire-Lincolnshire SE1	0.64	0.65	0.34	0.32	0.40	0.63
Yorkshire-Lincolnshire SE2	0.79	0.75	0.61	0.50	0.68	0.75
Yorkshire-Lincolnshire SE3-5	0.74	0.69	0.51	0.33	0.50	0.60
Mean correlation coefficient	0.712	0.683	0.529	0.336	0.567	0.659

These correlation coefficients show a clear pattern that the best correlations between forecast and actual admissions were consistently obtained for higher levels of aggregation (and those regions with larger numbers of emergency admissions). The poorest correlations were obtained for the 0-14 age group.

In most models, the temperature variables made a very small contribution to forecasts. In several models no weather variables were statistically significant and, therefore, were excluded through the step-wise process. Weather variables typically contributed less than 1.0% to explained total variation in the models. In the best fitting models (Southeast), the weather variables explained an average 0.64%, 0.50% and 0.89% of the total variation in admissions for socioeconomic group 1, socioeconomic group 2 and socioeconomic group 3-5 respectively. Weather variables, where retained, explained even less in poorer fitting models. For the Southeast models, an average of 3 out of 21 weather variables were retained in the socioeconomic groups 1 and 2, and 4 out of 42 for socioeconomic group 3-5.

Because these weather variables were calculated for between 2-4 isothermal regions within each socioeconomic group, it was not surprising that any single weather variable had little effect on hospital admissions for the region as a whole. As noted above, the more efficient approach would be to construct the weather variables for each socioeconomic area.

The overarching factors in all models were the intercept, and dummy variables for month and day of week.

Diagnostics

The output from S-Plus excluded any diagnostic tests (other than an F-test and R^2). The goodness-of-fit should, as a minimum, be measured by an adjusted R^2, an R_D^2 around the D differences of the data, or t-step ahead prediction errors. How well the models forecast actual admissions for the winter of 2002/03 is currently underway and will be reported separately. In time-series data, the primary diagnostic tests should include tests for autocorrelated errors, heteroskedasticity (non-constant variance) and tests for normality of residuals (zero mean, normal distribution, no excess kurtosis and skewness).

In data containing trends (i.e. time-series), it is not uncommon for errors[6] from a regression model to be related to previous errors; that is errors show a pattern over time. This could be due to excluded variables, incorrect functional form, or lags in the dependant variable not captured by the explanatory variables (e.g. common trends in the data or observations do not represent independent draws from the population). The parameter estimates from a model containing autocorrelated errors will be unbiased, but the standard errors of those parameters, and the model (R^2) will be understated leading to artificially high t-test and F-test statistics.

Similarly, in data containing trends, the variance of the residual might change with different levels of admissions (e.g. hospitals with large numbers of admissions might show more or less variance in admissions compared with smaller hospitals; or, the same hospital might have increasing variance in admission patterns over time). This heteroskedasticity can be caused by misspecification (e.g. omitted variable bias, unnecessary included variables, or incorrect functional form) and results in a non-constant variance of the error. The consequences of this are similar to those for autocorrelated errors – the parameter estimates are unbiased but the standard errors of those parameter estimates are understated leading to artificially high t-test statistics.

[6] The term "error" is used here as the difference between the estimated (predicted) level of admissions and the observed actual admissions. This is interchangeable with the term "residual" and forecast error.

Errors that are not normally distributed also bias standard errors of parameter estimates. Non-normal errors can result from misspecification – especially incorrect functional form (e.g. untransformed data when a log transformation should be used), or outliers (e.g. some periods with extremely high (or low) admissions). The approach taken to develop the models excluded exploring the most appropriate functional form. Without exploring the data with respect to functional form, retaining data in an untransformed state imposed linear restrictions *a priori* potentially causing misspecification bias.

Overall, without diagnostic tests being conducted during the development of the models, it is possible (and indeed probable) that models might suffer undue autocorrelation heteroskedasticity, and/or misspecification bias. This potential problem is compounded by the use of the step-wise procedure used for estimating the models. The step-wise procedure included/excluded variables at a threshold 5% level of significance. However, if the parameter standard errors are biased, then the variables will be incorrectly retained or excluded.

In the remainder of this section some basic tests for autocorrelation, heteroskedasticity, normality of errors and trends in the errors, are undertaken. These tests were undertaken for two "good" models (Southeast SE1 total and medical) and two "poor" models (Northern SE1 total and medical) for the data period used to estimate the coefficients for the Met Office models (i.e. April 1995 to March 1999, daily observations, n=1461).[7]

Tests for autocorrelation

Autocorrelation was tested by YHEC by running ordinary least squares regressions on the errors from the Southampton models.[8] The error variables (i.e. residuals from the regressions) were the dependant variable, and 7 lags were used as the explanatory variables.[9] Significant coefficients on these lags indicate significant autocorrelation at the corresponding lag.

In addition, the errors from the Southampton models were summed for each month to test for autocorrelation at monthly frequencies. Lags of nine months were used. It is possible that errors will be correlated at higher frequencies of one year. However, several years of data will be required to test at the annual frequency.

7 The models with the best and worst correlations between forecast and actual admissions were chosen as the "good" and "poor" models.

8 This is a quasi-Lagrange Multiplier technique.

9 The use of 7 lags is relatively small considering errors are likely to be correlated at an annual frequency (i.e. requiring 365 lags). The ideal number of lags to include is calculated, following Schwert (1989), as $l = \text{int}\{12(T/100)^{1/4}\}$ where T is the number of observations. For 1,461 observations, the number of lags = 23, and for four years on monthly observations, the number of lags = 9.

Tests for heteroskedasticity

Tests for heteroskedasticity were conducted by YHEC Ltd by summing the squares of errors from the first third of the observations and the last third (n=487). The ratio of the first to the last thirds is tested against an *F*-distribution with 487 degrees of freedom. Values close to unity indicate homoskedasticity; values significantly less than unity indicate greater variance in the latter third of the observations, and values significantly greater than unity indicate greater variance in the first third of the observations.

Tests for normality

Normality of the errors is tested with the Kolmogorov-Smirnov test. This test is based on mean and variance of the series being tested and is compared against expected values from a normal distribution. In addition, tests for skewness (symmetry of the distribution) and kurtosis (clustering around the central point of the distribution) and their standard errors, are reported. A skewness value of zero indicates symmetry; significant positive values indicate a long tail to the right (i.e. significant positive errors from overestimates of admissions); and significant negative values indicate a long tail to the left (i.e. significant negative errors from underestimates of admissions). Positive kurtosis indicates that the observations cluster more and have longer tails than those in the normal distribution and negative kurtosis indicates the observations cluster less and have shorter tails.

Tests for trends

A simple test for trends in the errors was undertaken by YHEC by running a linear regression with the Southampton error as the dependent variable and using the date as an explanatory variable. A significant coefficient on this variable indicates a trend in the data.[10] This, in turn, suggests misspecification bias, biased *t*-tests and *F*-tests resulting in potentially wrong selection of variables in the step-wise procedure, inflated R^2 values, and underperformance with respect to forecasts.

Tests for independence of models

This is a minor issue. However, it is possible that emergency admissions in one area might affect emergency admissions in a neighbouring area and, therefore, forecast errors might show similar patterns. To test for this, correlation coefficients were estimated using the forecast errors from total admissions in the two regions (Southeast Region and Northern Region).

---

[10] This is a relatively crude test for a trend but sufficient for the purposes of this review; other more sophisticated tests exist, such as unit root tests for stationarity (for example, see Harris R, 1995).

Results of diagnostic tests undertaken by YHEC

There was significant positive autocorrelation of errors in all models, and mainly at lags of 1, 6 and 7 days (Table 2.4)[11]. Results from using the sum of monthly errors shows that this positive autocorrelation was of sufficient magnitude to persist for one-month in all models, and continued to be significant in the model for total admissions in the Southeast region for three months. This positive autocorrelation will underestimate standard errors inflating t-statistics (and, therefore, incorrectly retaining variables that are not significant in the step-wise procedure), and exaggerate R^2 values. To determine if there was a pattern with monthly errors the daily error models were extended by including lags of 31 days (Appendix B). The pattern appears that errors are correlated at days 1 and 6 in all models, and significant positive autocorrelation remained at days 21, 26 and 30 in the Southeast models.

[11] Positive autocorrelation is the error in the current period is directly related to errors in previous periods. In contrast, negative autocorrelation is where current and previous errors are inversely related.

Table 2.4: YHEC diagnostic tests

	Southeast region SE1		Northern region SE1	
	Total admissions	**Medical admissions**	**Total admissions**	**Medical admissions**
Mean daily admissions[a]	577.5	362.1	70.0	45.5
Autocorrelation tests – *daily admission errors*				
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Constant	0.026	0.017	-0.002	0.006
Lag 1	0.198***	0.231***	0.063***	0.053**
Lag 2	0.124***	0.132***	0.016	0.050**
Lag 3	0.045*	0.082***	0.001	0.007
Lag 4	0.090***	0.083***	0.029	0.076***
Lag 5	0.015	0.002	0.032	0.033
Lag 6	0.084**	0.083***	0.063***	0.083***
Lag 7	0.065*	0.048*	0.085***	0.092***
Autocorrelation tests – *monthly admission errors*				
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Constant	29.113	24.075	1.416	1.946
Lag 1	0.652***	0.466***	0.380**	0.363**
Lag 2	-0.331**	-0.129	0.148	0.174
Lag 3	0.312**	0.200	-0.002	0.037
Heteroskedasticity test				
F(n,n) test [b]	1.105	0.949	0.893	0.784***
Normality tests				
Mean	0.0001	-0.0001	0.0004	-0.0001
Variance	1271.33	794.822	83.644	55.437
Mnimum	-208.30	-167.17	-28.95	-24.35
Min/mean adms (%)	48.5%	46.2%	41.4%	53.5%
Maximum	184.86	150.14	40.87	33.47
Max/mean adms (%)	32.0%	41.5%	58.4%	73.6%
K-S	1.054	1.082	1.290*	0.840
Skewness	-0.122	0.006	0.185**	0.226**
Kurtosis	2.787***	2.606***	0.415**	0.207
Trend test				
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**
Constant	-2765.96***	-2329.84***	-62.53	-186.00***
Trend	2.115E-07***	1.78E-07***	4.781E-09	1.422E-08***

Notes:

*, **, *** indicate significance at the 10%, 5% and 1% levels respectively

a. Mean admissions for the four-year period, April 1995 to March 1999

b. n=487; 5% critical values = 1.146, 0.854; 1% critical value = 1.122, 0.878

Tests for heteroskedasticity show that variance was constant in three of the four models tested, but the variance in medical admissions in the Northern region shows variance increased in the latter 16 months compared with the first 16 months. To correct for this requires either transforming the data (e.g. by logs) or using different estimation techniques (e.g. using an autoregressive conditional heteroskedasticity (ARCH) model).

The tests for normality show that the mean error in all models was zero, and the two *good* models (Southeast region SE1) had the largest error variance. The range of error values was greatest for the Southeast SE1 models, ranging from –208 to 185. However, variance and range, must be viewed relative to the mean admissions over the same period.

Based on the minimum and maximum errors as a percentage of mean admissions, the Southeast model under forecast emergency admissions by as much as 49% and over forecast by 32%. At the other end of the model performance spectrum, the Northern medical admission model under forecast by as much as 54% and over forecast by as much as 74%. These ranges in forecast errors could be due to some outliers. The Northern region models show significant positive skewness, indicating there were some large extreme observations. In contrast, the models for the Southeast show significant kurtosis; this indicates a tendency for observations to cluster around the mean.

These tests are of limited value without testing for trends in the data. The trend tests show significant trends in three of the four models. Therefore, without removing the trends all *t*-tests and *F*-tests are invalid. Because variable selection was based on *t*-tests in the step-wise procedure, the variables selected might not be the best. In addition, omitting a trend factor from forecasts beyond the period of 1995-1999 will, as the forecast horizon increases, tend to increasingly under-forecast admissions in subsequent periods.

Tests for independence showed a significant correlation coefficient of 0.179 ($p<0.0001$) for total forecast errors between the two regions. This is a relatively large correlation given that these regions are not neighbours. Higher correlations are expected for contiguous regions. This correlation suggests that factors influencing admissions in one region affect the other. The direction of causation is unknown without undertaking substantial additional analyses, but it is possible that causality is bidirectional.[12]

[12] For examples, see: Granger CWJ (1995) or Granger CWJ, Huang BN, Yang CW (2000).

Section 3: Use of the Models by the Met Office

3.1 USE OF THE MODELS BY THE MET OFFICE

Aggregation from postcodes to hospital

The coefficients from the models are used by the Met Office to forecast emergency admissions for each set of postcodes. "Currently, the performance of the model(s) is poor at the hospital level" (Doc 2). HES data for 2000/01 was used to estimate hospital catchment areas. The number of admissions from each postcode district to each hospital were counted, excluding postcode districts outside England, and postcode districts where the numbers were very small were merged with adjacent postcode districts. Then, for admissions from each postcode district, the percentage going to each hospital was calculated.

This approach of aggregation is reasonable. For socioeconomic groups 3-5, more efficient use of the temperature information could be made if temperature variables were constructed for each of the three socioeconomic groups within each PMS region (rather than for each of the isothermal sectors within socioeconomic group 3, then group 4, then group 5). This would provide fewer, and more relevant, coefficients to work with. The forecasting performance of the models might then be improved, and the process of aggregation, and the need for error tracking (see below) might be reduced from using a higher level of aggregation at the outset.

The objective was to forecast geography-based emergency admissions, with the emphasis on geography, rather than the hospital as the unit of analysis. The hospital as the unit of analysis appeals but because hospitals might be opened, closed or change function (e.g. from long-term care to acute care) over time, modelling admissions based on geography is relevant and the best approach.

Error tracking

Following aggregation of forecasts for postcodes to hospitals, error tracking was applied to correct for under or over-forecasts. Error tracking is a means of adjusting the forecast for each hospital such that it reflects recent real-time data and can, therefore, only be used for those hospitals which submit daily data. The error correction for each hospital was calculated as the average forecast errors over the last 28 days. This approach does lead to improved forecasts. Further improvement can be achieved should the forecast errors be weighted such that the most recent errors carry a greater weight, and errors from 28 days ago (or longer) have diminishing importance.

There are alternative error correction approaches which can be used. Most must be employed during the parameter estimation phase of modelling rather than following the aggregation phase.

Part of the need for error tracking arises from the presence of long-run trends in the data, and because the period of data used for generating the coefficients relates to a period ending four years ago. Systematic under-forecasts are most likely to be the result of omitting long-run trends.

The models estimated are static models – that is, short-run dynamics are not included. Short-run dynamics are an essential component of forecasts of many economic series as well as weather and in this case admissions to hospital. Including short-run dynamics might lead to improved forecasts. Several approaches to modelling time-series data include long-run trends, short-run dynamics and an error-correction mechanism.[13] Future analyses could be improved by including both long-run trends and short-run dynamics.

13 See Harris R (1995) or Banerjee A, Dolado J, Galbraith J, Hendry (1993).

Section 4: Discussion

4.1 DISCUSSION

The models developed by Southampton were a major undertaking and form a good basis for future development. Many aspects of the model development are commendable. For example, separating PMS regions into socioeconomic clusters based on Census data was an efficient approach to incorporate cross-sectional economic information into time-series analysis. The models developed for total admissions provided the best fit between actual and forecast admissions with a mean goodness-of-fit of 71%. Models for the 0-14 year age-group provided the worst fit with mean goodness-of-fit of 34%.

The overarching factors in explaining and forecasting emergency admissions to hospitals were the day of the week and month of the year. Temperature appears to have only a very small effect, and only in some areas, on admissions. This is possibly due to little variation in temperature from day-to-day, and in practice temperature may not have any real effect on being admitted to hospital in the UK. A more focussed study of occasions when the temperature does significantly change (i.e. modelling volatility), may lead to different conclusions.

One notable feature was model forecasting accuracy which was better with higher levels of aggregation. That is, when admissions for smaller groups were forecast, forecast accuracy was not so good. This might be due to random variation within small samples, or volatility to unobserved factors. It is modelling this volatility that will produce the greatest benefit with respect to modelling emergency admissions.

The models developed by Southampton are a first step into forecasting hospital admissions for England's total population. However, there are limitations with the models developed. Some of these are relatively minor and pertain to the data used, some are statistical issues on how the data was used, and some are conceptual issues. All of these limitations need to be addressed in future model-development work on forecasting hospital admissions.

The major limitations were:

- Socioeconomic clusters were based on the 1991 census; these groupings are likely to have changed over the last 12 years;
- The models were constructed with hospital and temperature data from April 1995 to March 1999; the coefficients from that period were used to generate forecasts in 2003. Forecasts made four years into the future based on four previous years of data

are unlikely to be robust. (This is the subject of a separate report by YHEC reviewing the forecasts for the winter of 2002/03);

- Trends in hospital admissions and population growth were not considered. This results in poor forecasts and invalid statistical tests (e.g. *t*-tests and *F*-tests).

It is possible systematic errors exist. Tests were conducted on autocorrelation of errors with 31 daily lags and three-month lags, but it is possible a wider pattern exists in the structure of the error. For example, there might be annual patterns or cycles that were not detected. In addition, it is possible there might be some pattern that exists across models – that is, each model was treated as independent, whereas, admissions in one region might be associated (either directly or inversely) with admissions in contiguous regions. This is a hypothesis that requires testing in future research.

The temperature variables were numerous, and many of these could be made redundant through aggregation at the conceptual stage of the project rather than following the analysis. The unit of analysis was, not the postcode, but 3 socioeconomic clusters within each of 9 PMS regions. Therefore, temperature data could be simplified into the same regions rather than the 2-4 isothermal sectors within each socioeconomic group.

The step-wise procedure for selecting variables and estimating coefficients is flawed; the t-test on which entry/retention/exclusion of variables was based is biased. Analysis of the error data showed substantial autocorrelation as well as showing the data contained time trends. This gives rise to model misspecification (i.e. exclusion of relevant variables, and common factors).

The error-tracking approach employed by the Met Office is not the best method to adopt to improve this type of forecasts. Models that take into account underlying trends, and use temperature variables for each socioeconomic group as a whole might lead to improved forecasts with less need to correct for forecast errors. Similarly, the models developed were static; no account of dynamics of bed numbers changing.

Bed managers typically make subjective forecasts based on the number of admissions for the same time last year, for the day of the week, plus a correction factor for increases in population. It is expected that the Met Office forecasts using the Southampton models will provide more accurate forecasts than hospital bed manager's subjective forecasts.

Recommendations for future modelling of admissions

There are some important aspects that need to be refined and addressed in future development of the models. These include:

- Updating socioeconomic cluster boundaries based on the 2001 Census data;

- Aggregating temperature variables to match with the boundaries of the socioeconomic groups;
- Including trends in the analysis;
- Use all variables in the analysis (rather than a step-wise approach) or control for autocorrelation;
- Including dynamics for population and hospital changes over time;
- Test for independence of the models by checking correlation of errors and lags of errors across contiguous regions.

References

Banerjee A, Dolado J, Galbraith J, et al. (1993) *Co-integration, Error-correction and the Econometric Analysis of Non-Stationary Data.* Advanced Texts in Econometrics, Oxford University press, Oxford.

Bennett J, Pitman R, Jarman B, et al. (2001) A study of the variation in tuberculosis incidence and possible influential variables in Manchester, Liverpool, Birmingham and Cardiff in 1991-1995. *International Journal of Tuberculosis and Lung Disease.* 5(2):158-163.

Duffield JS, Craig K, Plant WD. (1997) Patterns in acute referral to hospital. *Scottish Medical Journal*. 42(4):105-107.

Fleming DM. (2000) The contribution of influenza to combined acute respiratory infections, hospital admissions, and deaths in winter. *Communicable Disease and Public Health.* 3(1), 32-38.

Fleming DM, Cross KW. (1993) Respiratory syncytial virus or influenza? *The Lancet.* 342, 1507-1510.

Granger CWJ. (1995) Causality in the long-run. *Econometric Theory*. 11(3): 530-36.

Granger CWJ, Huang BN, Yang CW. (2000) A Bivariate Causality between Stock Prices and Exchange Rates: Evidence from Recent Asian Flu. *Quarterly Review of Economics and Finance.* 40(3): 337-54.

Harris R. (1995) *Using cointegration analysis in econometric modelling.* Prentice Hall, London.

Jarman B. (1984) Underprivileged areas: validation and distribution scores. *British Medical Journal.* 289: 1587-1592.

Kuh D, Stirling S. (1995) Socioeconomic variation in admission for diseases of female genital system and breast in a national cohort aged 15-43. *British Medical Journal.* 311(7009):840-843.

Nguyen-Van-Tam JS, Brockway CR, Pearson JCG, et al. (2001) Excess hospital admissions for pneumonia and influenza in persons ≥ 65 years associated with influenza epidemics in three English health districts: 1987-95. *Epidemiology and Infection.* 126, 71-79.

Reid FD, Cook DG, Majeed A. (1999) Explaining variation in hospital admission rates between general practices: cross sectional study. *British Medical Journal.* 319(7202):98-103.

Scuffham, P. (2003) Estimating Influenza-Related Hospital Admissions in Children and Adults: A time-series analysis. *Disease Management and Health Outcomes.* 11:4;259-269.

Schwert GW. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics.* 7 pp 147-59.

Simonsen L, Fukuda K, Schonberger LB, Cox NJ. (2000) The impact of influenza epidemics on hospitalizations. *Journal of Infectious Diseases.* 181(3), 831-837.

APPENDIX A

Example of the List of Variables Used for One PMS Region

A.1 EXAMPLE OF THE LIST OF VARIABLES USED FOR ONE PMS REGION

Socioeconomic group 1	Socioeconomic group 2	Socioeconomic groups 3-5	
ITSSEC11_maxt2	ITSSEC12_maxt2	ITSSEC13_maxt2	ITSSEC13_MinMax07
ITSSEC11_maxt7	ITSSEC12_maxt7	ITSSEC13_maxt7	ITSSEC13_MinMax10
ITSSEC11_maxt10	ITSSEC12_maxt10	ITSSEC13_maxt10	ITSSEC13_MinMax15
ITSSEC11_maxt15	ITSSEC12_maxt15	ITSSEC13_maxt15	ITSSEC23_MinMax07
ITSSEC21_maxt2	ITSSEC22_maxt2	ITSSEC23_maxt2	ITSSEC23_MinMax10
ITSSEC21_maxt7	ITSSEC22_maxt7	ITSSEC23_maxt7	ITSSEC23_MinMax15
ITSSEC21_maxt10	ITSSEC22_maxt10	ITSSEC23_maxt10	ITSSEC33_MinMax07
ITSSEC21_maxt15	ITSSEC22_maxt15	ITSSEC23_maxt15	ITSSEC33_MinMax10
ITSSEC31_maxt2	ITSSEC32_maxt2	ITSSEC33_maxt2	ITSSEC33_MinMax15
ITSSEC31_maxt7	ITSSEC32_maxt7	ITSSEC33_maxt7	ITSSEC24_MinMax07
ITSSEC31_maxt10	ITSSEC32_maxt10	ITSSEC33_maxt10	ITSSEC24_MinMax10
ITSSEC31_maxt15	ITSSEC32_maxt15	ITSSEC33_maxt15	ITSSEC24_MinMax15
ITSSEC11_MinMax07	ITSSEC12_MinMax07	ITSSEC24_maxt2	ITSSEC34_MinMax07
ITSSEC11_MinMax10	ITSSEC12_MinMax10	ITSSEC24_maxt7	ITSSEC34_MinMax10
ITSSEC11_MinMax15	ITSSEC12_MinMax15	ITSSEC24_maxt10	ITSSEC34_MinMax15
ITSSEC21_MinMax07	ITSSEC22_MinMax07	ITSSEC24_maxt15	ITSSEC35_MinMax07
ITSSEC21_MinMax10	ITSSEC22_MinMax10	ITSSEC34_maxt2	ITSSEC35_MinMax10
ITSSEC21_MinMax15	ITSSEC22_MinMax15	ITSSEC34_maxt7	ITSSEC35_MinMax15
ITSSEC31_MinMax07	ITSSEC32_MinMax07	ITSSEC34_maxt10	
ITSSEC31_MinMax10	ITSSEC32_MinMax10	ITSSEC34_maxt15	
ITSSEC31_MinMax15	ITSSEC32_MinMax15	ITSSEC35_maxt2	
		ITSSEC35_maxt7	
		ITSSEC35_maxt10	
		ITSSEC35_maxt15	

Notes:

- The first digit following ITSSEC refers to the isothermal sector within that socioeconomic group;
- Maxt refers to the maximum temperature at days 2, 5, 10 or 15 previously;
- MinMax refers to the lowest maximum temperature recorded within the respective isothermal sector within the socioeconomic group in the previous 7, 10 and 15 days.

APPENDIX B

Tests for Autocorrelation of Errors With Lags of 31 Days

B.1 TESTS FOR AUTOCORRELATION OF ERRORS WITH LAGS OF 31 DAYS

	Southeast region		Northern region	
	Total admissions	Medical admissions	Total admissions	Medical admissions
	Coefficient	Coefficient	Coefficient	Coefficient
Constant	0.118	0.087	-0.008	0.010
Lag 1	0.188***	0.228***	0.056**	0.052*
Lag 2	0.116***	0.123***	0.018	0.038
Lag 3	0.039	0.077***	-0.004	0.001
Lag 4	0.084***	0.080***	0.025	0.069***
Lag 5	-0.004	-0.012	0.033	0.021
Lag 6	0.076***	0.070**	0.063**	0.080***
Lag 7	0.047*	0.027	0.082***	0.083***
Lag 8	0.042	0.064**	0.014	0.000
Lag 9	0.008	-0.000	0.043	0.071***
Lag 10	0.004	0.004	0.005	-0.013
Lag 11	0.019	0.018	0.004	0.031
Lag 12	0.035	0.040	-0.024	-0.023
Lag 13	-0.044	-0.022	-0.013	-0.013
Lag 14	0.019	0.035	0.005	0.023
Lag 15	0.000	-0.034	0.007	0.031
Lag 16	-0.007	-0.027	0.044	0.011
Lag 17	0.018	0.004	0.027	0.032
Lag 18	0.004	-0.003	0.004	0.004
Lag 19	-0.002	-0.009	0.003	-0.002
Lag 20	0.002	0.017	0.025	0.030
Lag 21	0.078***	0.073***	-0.003	-0.005
Lag 22	-0.018	-0.063**	0.010	0.007
Lag 23	0.008	0.019	-0.001	-0.018
Lag 24	-0.041	-0.013	-0.048*	-0.012
Lag 25	0.047*	0.040	0.005	0.016
Lag 26	0.011	-0.018	0.030	0.011
Lag 27	-0.061**	-0.050*	0.003	0.018
Lag 28	-0.016	-0.003	-0.003	-0.022
Lag 29	0.005	-0.014	0.044*	0.033
Lag 30	0.087***	0.067**	-0.008	-0.006
Lag 31	0.010	0.035	0.033	0.025
Adjusted R^2	0.154	0.197	0.013	0.032

*, **, *** indicate significance at the 10%, 5% and 1% levels respectively